ISSN: 2366-7109

WORKING PAPERS OF AGRICULTURAL POLICY

AGRICULTURAL POLICY
WORKING PAPER SERIES
WP2018-03

Applying Meta-modeling for extended CGE-modeling: Sampling techniques and potential application

Ding Jin
Johannes Hedtrich
Chrisitan Henning

Department of Agricultural Economics
University of Kiel

**The Agricultural Working Paper Series is published by the Chair of Agricultural Policy at the University of Kiel.
The authors take the full responsibility for the content.**

Ding Jin
Johannes Hedtrich
Chrisitan Henning

Applying Meta-modeling for extended CGE-modeling:
Sampling techniques
and potential application

Department of Agricultural Economics
University of Kiel

Kiel, 2018
WP2018-03
http://www.agrarpol.uni-kiel.de/de/publikationen/working-papers-of-agricultural-policy

About the authors:
Christian Henning is professor and chair of agricultural economics, University of Kiel (Germany). He studied economics, agricultural economics, sociology, political science and mathematics. He earned a PhD in economics, in agricultural economics and in political science. His main areas of specialization are applied political economy and applied economic modeling of agricultural policies.
Ding Jin is a doctoral candidate under the supervision of Professor Dr. Christian Henning at the University of Kiel. He studied quantitative economics and earned a master degree in quantitive economics.
Johannes Hedtrich is a doctoral candidate under the supervision of Professor Dr. Christian Henning at the University of Kiel. He studied computer science and earned a master degree in computer science.
Corresponding author: djin@ae.uni-kiel.de

Abstract

Apart from the computational time and expenses of the CGE model, the discussion of elasticity parameter estimation and various closure rules as well as the difficulty of combining the results with other analysis approaches always poses obstacles ahead of us, therefore we are motivated to apply the Bayesian model selection method and the meta-modeling technique in order to tackle these problems from a new perspective in the framework of the Senegal-CGE model and even compare the CGE models. The meta-modeling technique can be deemed as an intermediate step towards the application of Bayesian model selection method because the CGE models cannot be directly integrated into the method. The meta-modeling technique, whose three essential components are the simulation models, the meta-models and the design of experimetns, aims at generating valid and simplified approximation models of the simulation models and gives us the opportunity of combining the CGE models with the Bayesian model selection method. The purpose of this paper is to demonstrate the meta-modeling techinique, test the performace of the meta-models generated by it and analyze whether various combinations of elasticity parameters affect the outputs of the CGE models which are quantified by the marginal effects. Our findings show that the produced meta-models possess a decent prediction capacity but we have not detected significant variability of the marginal effects within each unique sector.

Keywords: Bayesian model selection; CGE modeling; Elasticities; Closure Rules; Meta-modeling; Meta-models; DOE
JEL classification:

1 Introduction

CGE-applications are workhorse models in applied economic policy analysis, i.e. the development economic literature or modeling climate and energy policies. However, beyond its prominent application CGE-model approaches are also heavily criticized. On the one hand, while the general equilibrium model has the advantages in terms of internal consistency and allowing for clearer identification of causality, the application of a CGE model requires simplifying assumptions that are open to challenge. Moreover, empirical results derived from the CGE-model application are very sensitive to specific model specifications, that are often only weakly empirically justified, e.g. assumed closure rules and assumed elasticity parameters. Thus, many results, e.g. growth-poverty linkages, that are derived from a CGE model are in fact plagued by high model uncertainty implying a limited potential to generate robust policy-relevant messages.

On the other hand, CGE-model approaches are often applied to provide scientific expertise to advise the government in political practice. Hence, it would be necessary to incorporate general equilibrium models into overall decision-making models. However, given the size and complexity of CGE-models integration of these approaches into an overall decision-making modeling approach is rather difficult and often numerically not tractable.

Therefore, in the context of such a situation, we suggest application of Bayesian model selection method which is applied to select the “correct” CGE model from the possible candidates on the basis of empirical data and expert data in order to tackle the problem of parameter estimation and closure rule assumptions. Due to the fact that Bayesian model selection method cannot work directly with the CGE models because they are too complex to be incorporated into the method, we are in need of simplified and valid surrogate models of the CGE models to be replaced in the implementation of the Bayesian model selection method. To fulfill this purpose, we apply the meta-modeling technique which can be viewed as an intermediate step to generate the approximation models of the CGE models. By means of the meta-modeling technique, we are able to produce surrogate models for all the possible CGE models determined by different behavioral parameters and closed by various closure rules and select the “correct” one using the Bayesian model selection method. In spite of the fact that this is not discussed in our paper, we still want to point that the meta-modeling technique enables us to generate simplified surrogate models of the CGE models and it gives us the chance to intergrate them further into the decision-making models as well.

Estimation of the behavioral parameters and assumptions of the closure rules can be thought of as a forward solution to the existing problems of CGE modeling. On the contrary, we would like to suggest what can be considered as a backward solution to the problem which we bypass the estimation of behavioral parameters and assumptions of closure rules, hypothesize that we have a databese of all the possible candidates and try to select the “correct” from them. In this paper, however, our main purpose is to demonstrate the meta-modeling methodology.

The paper is arranged as follows: section 2 provides a general literature review; section 3 gives the the description of the meta-modeling methodology as well as the building blocks of it; section 4 demonstrates the results and section 5 concludes the paper.

2 Literature Review

CGE-modeling is a common workhorse in development economics and policy analysis. It has been widely used to model climate and energy policies as well. (Bourguignon (2003); Lofgren et al. (2002); Fan (2008)). However, in spite of its prominent applications, CGE-modeling has been heavily criticized because empirical results derived from CGE-models are very sensitive to specific model specifications, that are often only weakly justified, e.g. assumed closure rules and assumed elasticity parameters. (Lofgren and Robinson (2008); Hazledine et al. (1992); Arndt et al. (2002)) Thus, many results, e.g. growth-poverty linkages, that are derived from a CGE model are in fact plagued by high model uncertainty implying a limited potential to generate robust policy-relevant messages. (Lofgren and Robinson (2008)). Thus, we suggest the application of Bayesian model selection method which aims at selecing the “right” model with all the information available. As the CGE models are too complex to be directly incorporated into the method, we need simplified and valid surrogate models of the CGE models which can be generated via the met-modeling technique. Meta-modeling technique has been extensively used in field such as engineering, natural science, production design and etc. (Srivastava et al. (2004); Noordegraaf et al. (2003); Kleijnen and Standridge (1988)) The basic approach is to construct approximation models of the simulation models in order to generate surrogate models that are accurate and reliable enough to replace the original ones with the purpose of understanding the simulation models better and combining the simulation models with other analysis methods.(Kleijnen and Sargent (2000); Kleijnen (2008)) Building approximation models include two essential components: design of experiments and meta-models, the former is used to produce the simulation sample (Kleijnen et al. (2005); Giunta et al. (2003); Eriksson et al. (2000)) while the latter is used to determine the form of the surrogate models. (Simpson et al. (2001); Wang (2007)). By means of the meta-modeling technique, we are able to obtain surrogate models of the CGE models that can be used in the Bayseian model selection method. To the best of our knowledge, despite the rich applications of the meta-modeling technique in other areas, it has not been used in company with CGE models. Thus, this paper aims at demonstrating the meta-modeling technique and provide an extension of the application of CGE models.

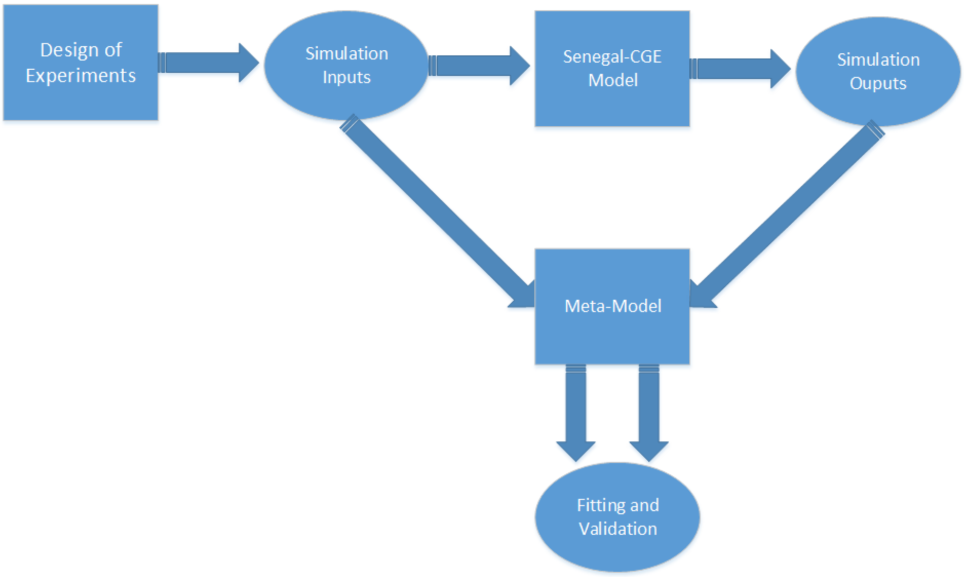


Figure 1: The Meta-modeling Flow.

3 Meta-modeling Methodology

The general meta-modeling methodology (Figure 1) can be described as folows:

1. The simulation model (the Senegal-CGE model) is treated as a black box and we assume a meta-model for it.

2. The design of experiments method is applied to generate the simulation inputs, denoted as x, and they are used to produce the simulation outputs, denoted as z.

3. The simulaton inputs and outputs are collected in order to fit and validate the meta-model. If the validation criteria are met, we can use the meta-model for other research purposes.

As we can readily see that there are three essential components of the meta-modeling technique which are the simulation model, the design of experiments method and the meta-models. These three components interact with each other in the implemetation of the meta-modeling methodology, therefore we need to pay attention to understanding the mechanism of them.

In addition, table 1 lists the detailed meta-modeling process which contains the following 7 steps (Barton (2015)). Alongside of the description of the steps, we will also introduce the important concepts of the method.

3.1 Meta-modeling Purposes

The general purpose of implementing meta-modeling technique is to generate simplified and validated meta-models for the simulation models. The term meta-models in our

Step	Activity
1	Meta-modeling Purposes
2	Inputs and Outputs Indentification
3	Meta-model Types
4	Experimental Design based on Meta-model Types
5	Conducting Simulation Runs and Collecting Outputs
6	Fitting and Validation of Meta-model Adequacy
7	Use Meta-model for Other Research Purposes

Table 1: Meta-modeling Process

context are equivalent to the surrogate models. There are several reasons that researchers apply this method to obtain meta-models:

1. The simplicity to construct and understand.
2. The need of less computational time and expenses.
3. The possibility of combining with other approaches.

The merits of meta-models mainly rest on the fact that the meta-models are by nature mathematical functions such as $z = \beta_0 + \beta_1 x1 + \beta_2 x2$. It's easy to understand that if we are able to generate such a mathematical function as a surrogate model for a complex simulation model, we can benefit greatly from using it.

In addition to the general purpose, as stated in section **??**, we are looking forward to applying the Bayseian model selection method to tackle the aforementioned CGE modeling problems as well as integrate the CGE models into the decision-making models, therefore, we are also in need of the meta-models to replace the simulation CGE models.

3.2 Inputs and Outputs Indentification

Regarding the inputs, the Senegal-CGE model contains a number of variables and for our current purpose, we take three categories of them into consideration (in total we have 20 variables of interests):

1. Policy indicators. The policy indicators are actually the targeted outcome of implementing a policy. For example, in our case they are the technical progress shocks in eight aggregated sectors: crop, export, livestock, fish, agribusiness, industry, service and public. We assume that the policy will generate the technical progress shocks in these sectors. As for now we don't focus on the policy-growth linkage [1], all of them are assumed to range from 1% to 10%.

[1] The policy-growth linkage refers to the idea that the implemetation of a certain policy will lead to some growth shocks in the CGE model such as technical progress shocks in our case.

2. Production elasticities (factor subsitution elasticities) in eight aggregated sectors: crop, export, livestock, fish, agribusiness, industry, service and public. All of them range from 1.5 to 6.

3. Trade elasticities (Armington transformation elasticities) in agricultural and nonagricultural sectors. The trade elasticities in agricultural sector range from 0.5 to 3.3 and the trade elasticities in nonagricultural sector range from 0.9 to 4.1.

The three components mentioned above are all simulations inputs, i.e. the x in mathematical terms. In other words, they will be generated by means of the design of experiments methods which we will introduce later. Moreover, we want to exphasize again that the elasticity parameters are generated not estimated which is exactly the difference of this approach. We treat the elasticity parameters as independent variables due to the reason that we assume that they play a role in affecting the outputs. We will explain more about this in the following sections. Besides, in terms of the values of all the inputs, they are specifically determined based on practical considerations.

Regarding the outputs, z, the Senegal-CGE model has seven outputs, which are: $z1$ (Small Household Income), $z2$ (Poverty Reduction Index), $z3$ (General Public Services), $Z4$ (Welfare of Agribusiness), $z5$ (Urban Consumer Welfare), $z6$ (Welfare of Agricultural Export Sectors), $z7$ (Environmental Protection). In this paper, we will analyze $z1$, $z2$, and $z5$ in order to test the performance of meta-models and the impacts of reduced-form meta-models.

In the rest of the paper, the inputs are denoted as x and the the outputs are denoted as z.

3.3 Meta-model Types

The meta-modeling technique includes three essential components: the simulation model, the meta-model, and the experimental design.(Kleijnen and Sargent (2000)) The meta-model is a mathematical approximation equation that we assume and use to approximate the Input/Output behavior of the simulation model (the Senegal-CGE model in this paper). In this section, we give an introduction of the concept meta-models.

Meta-models aim at approximating the Input/Output relationships of simulation models. The term meta-model was popularized and developed by Jack Kleijnen(Kleijnen (1975)), but the term and concept were both originated by Robert Blanning(Blanning (1974); Blanning (1975)). Meta-models are usually used to model the behavior of another model and they are also termed surrogate models or response surface models. In the history of meta-models, they are applied to approximate both the stochastic simulation and the deterministic simulation.

A meta-model is a mathematical function that takes some simulation model design parameters [2] as inputs and produces an approximation of simulation outputs.

There are many types of meta-models in the literature and for our current research purpose, we will focus on two types of them, which are the lower-order polynomial and kriging meta-models.

3.3.1 Lower-Order Polynomial Metamodel

Lower-order polynomial meta-models are originally developed for the analysis of physical experiments(Box and Wilson (1951)) and they have been used effectively for building approximations in a variety of applications. There are different forms of this type but the most commonly used forms are first-order polynomial and second-order polynomial meta-models.

A second-order polynomial meta-model has the functional form:

$$z = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_i \sum_j \beta_{ij} x_i x_j + \epsilon, \tag{1}$$

where x_i and x_j are the model design parameters and z is the simulation outputs. Moreover, $\beta' s$ are the corresponding coefficients and ϵ is the error term which is often assumed to be a white noise process.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_i \sum_j \beta_{ij} x_i x_j + \epsilon \tag{2}$$

The corresponding coefficients $\beta' s$ are estimated using the ordinary least-squares regression and the estimates are computed as follows:

$$\hat{\beta} = [X'X]^{-1} X'z, \tag{3}$$

where X is the model design matrix and z is the simulation output. We can also perform other standard statistical analysis of the estimates.

Lower-order polynomial meta-models are attractive because they are easy to construct, understand and analyze. Besides, they work well in modeling local and linear behavior of the simulation model but if the simulation model is nonlinear or irregular, they might fail in approximating the behavior and we must resort to other meta-model types.

3.3.2 Kriging Metamodel

Kriging meta-models are originally developed for applications in geostatistics (Cressie and Chan (1989)), a kriging model postulates a combination of a polynomial model and

[2]Model design parameter is a term in the field of meta-modeling. It denotes the simulation input.

departures of the form:

$$z = \sum_{i=1}^{k} \beta_i f_i(x) + P(x), \tag{4}$$

where $f_i(x)$ is the polynomial model and $P(x)$ is assumed to be a realization of a stochastic process with mean zero and spatial correlation function given by:

$$Cov[P(x_i), P(x_j)] = \sigma^2 R(x_i, x_j), \tag{5}$$

where σ^2 is the variance of this process and R is assumed to be the correlation function of this process. A variety of correlation functions can be chosen, such as linear correlation function, exponential correlation function and Gaussian correlation function. Besides, as $P(x)$ is also assumed to be a stationary process so the covariances $R(x_i, x_j)$ are dependent only on the distance between the input combinations x_i and x_j. The corresponding coefficients are estimated using the maximum likelihood estimation method.

The Kriging meta-mdels are more flexible and can be used to model nonlinear or irregular behaviors of the simulation model.

The key idea is that we assume a specific meta-model for the simulation model which takes the simulation inputs, discussed in section 3.2, as independent variables and simulation ouputs, discussed in section 3.2, as dependent variables.

In the literaure, there are many other types of meta-models such as radial basis functions, neural networks, and regression trees. See Chen et al. (2006) for a review. The choice of suitable meta-model types relies on the problem setting we encounter and other practical concerns. However, due to the simplicity of construction and comprehension, lower-order polynomial meta-models are always a good place to start.

3.4 Experimental Design based on Meta-model Types

Design of experiments (Eriksson et al. (2000)), or DOE for short, is a sampling technique which we can apply to sample the model design space in order to generate the simulation sample. For example, we have k quantitative design parameters and each of them has n different values, which means that if we want to run all the possible scenarios, we would end up with n^k simulation runs and it could probably be a number that we are not able to handle. Therefore, we need a technique with which we can generate a workable sample while at the same time this sample must possess desirable properties and enough information for the follow-up analysis.

There is a large number of experimental designs in the literature, but for our current purpose we will discuss two types of DOE, the Central Composite Design and the Latin Hypercube Sample Design.

3.4.1 Central Composite Design

The Central Composite Design or CCD is a classical fractional factorial experimental design which spreads the sample points at three different places of the design space: (i) the vertices of the design space; (ii) the center of the design space; (iii) the star points which are placed along the axes but outside the design space (Giunta et al. (2003)).

A two-variable CCD contains the following sample points:

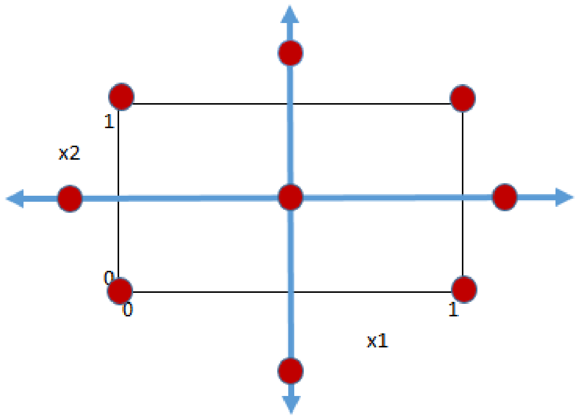


Figure 2: A Central Composite Design for $n = 2$.

The central composite design guarantees that the estimates of the coefficients of a second-order polynomial metamodel are unbiased. The number of sample points of a CCD follows the formula $2^n + 2n + C_0$, where n is the number of variables, 2^n is the number of sample points at the vertices, $2n$ is the number of star points and C_0 is the number of center points. In a central composite design, we can not control the number of sample points once n is fixed except that C_0 is an arbitrary number which we can alter. This means when the number of variables n grows, the sample points that we need to estimate the metamodel also increases exponentially.

3.4.2 Latin Hypercube Sample Design

Latin Hypercube Sample Design, or LHS for short, is a space-filling design which arranges the sample points as spread-out as possible across the design space in order to collect information inside the design space. Besides, LHS has another attribute that we can control the number of sample points based on practical concerns.

The Latin hypercube sample design works as follows (McKay et al. (1979); Stocki (2005)): suppose we have n variables and we need p sample points to fit our metamodel. Then the intervals of every variable are divided into p subintervals and one value is chosen out of every subinterval based on the probability density within that subinterval for each variable. Next, the p values of x_1 is paired randomly with the p values of x_2, then this established pair of x_1 and x_2 is again paired at random with the p values of x_3, and this process will be continued until the p n-tuplets are formed which are exactly the p sample points that we need for the simulation.

We can have a look at the following example with $n = 2$ and $p = 4$:

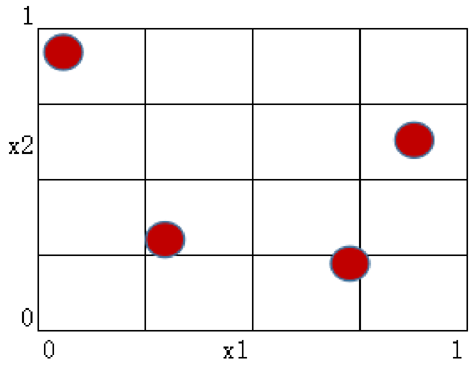


Figure 3: A Latin Hypercube Sample Design.

As a member of the space-filling design family, the latin hypercube design aims at placing the sample points as spread-out as possible across the design space. There are many criteria and optimality rules regarding generation of nice latin hypercube samples from which we list the following ones, such as (R package "lhs"):

1. Random LHS: draws a latin hypercube sample from a set of uniform distributions for use in creating a latin hypercube design. This sample is taken in a random manner without regard to optimization.

2. Improved LHS: draws a latin hypercube sample from a set of uniform distributions for use in creating a latin hypercube design. This sample is drawn based on the idea of optimizing the sample with respect to an optimum euclidean distance between design points.

3. Maximin LHS: draws a latin hypercube sample from a set of uniform distributions for use in creating a latin hypercube design. This sample is drawn based on the idea of optimizing the sample by maximizing the minimum distance between design points (maximin criteria).

4. Genetic LHS: draws a latin hypercube sample from a set of uniform distributions for use in creating a latin hypercube design. This sample is drawn based on the idea of optimizing the sample with respect to the S optimality criterion through a genetic type algorith. S optimality seeks to maximize the mean distance from each design point to all the other points in the design space, so the points are as spreat-out as possible.

5. Optimum LHS: draws a latin hypercube sample from a set of uniform distributions for use in creating a latin hypercube design. This sample is drawn using the Columnwise Pariwise (CP) algorithm to generate an optimal design with respect to the S optimality criterion.

The main message is that we apply this method to generate a sample of the simulation inputs discussed in section 3.2.

In this section, we have introduced two experimental designs: the central composite design and Latin hypercube design. There have been researches on the connection between the meta-models and experimental designs as well as the connection between the nature of simulations and experimental designs. (Simpson et al. (2001)) In this paper, because of the practical considerations such as computational time and the fact that we come across a deterministic Senegal-CGE model, LHS is a proper beginning point.

3.5 Conducting Simulation Runs and Collecting Outputs

The simulation model we use in this paper is a standard dynamic CGE model extended for Senegal. For a detailed description of CGE model, see Lofgren et al. (2002). The simulation inputs are generated using the design of experiments method and then passed into the simulation model to generate the simulation outputs.

3.6 Fitting and Validation of Meta-model Adequacy

We collect the simulation inputs using the DOE and the simulation outputs by running the simulation scenarios. Then we use them to fit the meta-model which we assume to has a second-order polynomial form. This process is not tricky because we use the standard OLS approach to estimate the coefficients. If the fitting is satisfactory, usually it is determined by the R^2_{adj}, we can move forward to validating the meta-model. There are various kinds of criteria which can determine the validation adequacy and we use the root mean squared prediction error(rmse) and the maximal absolute relative error(mare):

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{z}_i)^2} \tag{6}$$

$$mare = max|\frac{z_i - \hat{z}_i}{z_i}| \ \forall i, \tag{7}$$

where n is the number of observations, z_i is the simulation output and $\hat{z}_i$ is the corresponding output predicted by the meta-model.

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (8)$$

$$mare = max|\frac{y_i - \hat{y}_i}{y_i}| \quad \forall i \quad (9)$$

Besides, in order to be more certain of the validation adequacy, we apply both the in-sample validation and out-of-sample validation, therefore we need to split the sample into two parts. We have a sample with the size of 2000 and divide it into two subsamples, which are the training sample and the test sample, each with the size of 1000. The training sample is used to perform the in-sample validation while the test sample is used to perform the out-of-sample validation.

With respect to the in-sample validation, we use the cross-validation method. Firstly, the first observation is deleted and the meta-model is fitted using the rest 999 observations. Then the output of the deleted observation is predicted using the newly fitted meta-model. Secondly, the second observation is deleted and the meta-model is fitted using the rest 999 observations. Then the output of the deleted observation is predicted using the newly fitted metamodel. The same process will be applied to each observation of the training sample and 1000 predicted outputs will be computed. With respect to the out-of-sample validation, we use the training sample to fit the meta-model and then use the fitted meta-model to predict the outputs using the observations of the test sample. Thus we can calculate the two statistics for both the in-sample validation and out-of-sample validation.

3.7 Use Meta-model for Other Research Purposes

The CGE models are criticized for many reasons from which we will discuss two prominent ones: elasticity parameters and closure rules.

On the one hand, the CGE models are determined by the elasticity parameters in order to fit the data, namely, for each different set of elasticities, we have a completely different CGE model. In one case, some elasticity parameters cannot be estimated and researchers usually solve this problem by assuming them to be certain values, which leads to the fact that the results become very sensitive. In another case, the estimation and determination of some elasticities are not trivial because they require a large amount of econometric modeling, therefore in practice, most researchers recycle the estimates of others, though often modifying them for one reason or another or "......selecting these values by consulting econometric and other model-based studies......" (Lofgren and Robinson (2008)). Besides, empirically, researchers will perform the sensitivity analysis by altering the elasticities either to a high value or to a low value and comparing the result with that of the original value. This process, although it is useful for researcher to test the sensitivity of the elasticity assumptions, it still cannot provide valid guidance as to the rationality of the baseline elasticity assuptions. On the other hand, the CGE models aim to capture all the impacts, which might be distributed through the whole economy, of a shock such as tax rise or tariff cut. And the models are 'closed' by requiring that the total supply of an object must be equal to the total demand, with whatever adjustments needed to achieve it. In practice, however, there does not exist a true general model and the choice of closure rules is usually arbitrary, which will definitely affect the results. Both the elasticity parameters and the closure rules can to some extent impact final results. However, in practice, we usually we either don't have information on which of them are rational or complex econometrical models are needed to estimate the elasticity parameters, therefore in this paper, we would like to tackle this problem from a completely new perspective with the help of the Bayesian model selection method as well as the meta-modeling technique.

Theoretically, since the meta-model has fulfilled the validaton requirements, we can use it in lieu of the simulation model to perform the prediction and optimization tasks which is already is good result because using the meta-models is much easier faster and saves lots of time. However, our purpose of applying meta-modeling technique is to use the meta-models to assist the Bayesian model selection method, therefore, we need to create surrogate meta-models for all the possible CGE model candidates. This step is achieved by generating the reduced-form meta-models. The term reduced-form meta-model originates from the idea that we want to reduce the general meta-model, which takes policy indicators, production elasticities and trade elasticities as independent variables, to a more compact form, which takes only the policy indicators as independent variables while the production and trade elasticities are considered to be fixed values, so that the elasticities will be grouped into the corresponding coefficients. Apart from the reason that we need the reduced-form meta-models to apply the Bayseian model selction method, we also assume that the elasticity parameters play a role in affecting the outputs. Thus, by means of this approach we are able to detect this effect. The reason we generate them in this way is due to the fact that the CGE models are to some extent determined by the elasticity parameters, which in other words means that each unique combination of elasticity parameters is capable of representing a unique CGE model. Conventionally, a unique combination of elasticity parameters is either assumed or estimated. However, in line with our approach, we don't estimate or assume each unique combination of elasticity parameters with which we determine the CGE model, instead, we generate a huge sample of possible combinations using the design of experiments method which enables us to obtain a large database of the possible CGE model candidates. In this paper, we use the DOE method to generate a simulation sample with the size of 2000 and each of them contains a unique combination of elasticity parameters. Therefore, the simulation sample is equivalent to a database of 2000 CGE model candidates and by reducing the general meta-model we obtain the corresponding 2000 reduced-form meta-models. As to the closre rules, which will not be discussed in this paper though, we plan to run the 2000 simulation scenarios and 'close' the model under every possible combination of closures separately and generate the reduced-form meta-models affiliated to each combination of closure rules which will enlarge the database of possible CGE model candidates.

The reduced-form meta-models are generated as follows. Firstly, we fit and validate a general polynomial meta-model which takes policy indicators, production elasticities and trade elasticities as independent variables:

$$\begin{aligned}\hat{z} = \alpha_0 &+ \sum_i \alpha_i tp_i + \sum_j \beta_j \theta_j \\ &+ \sum_i \sum_{i'=i+1} \alpha_{ii'} tp_i tp_{i'} + \sum_i \sum_j \beta_{ij} tp_i \theta_j + \sum_j \sum_{j'=j+1} \gamma_{jj'} \theta_j \theta_{j'} \\ &+ \sum_i \alpha_{ii} tp_i^2 + \sum_j \gamma_{jj} \theta_j^2,\end{aligned} \tag{10}$$

where tp_i's are the technical progress shocks and θ's are the elasticity parameters. α's, β's and γ's are the corresponding coefficients.

The reduced-form meta-model takes only the policy indicators as independent variables while the production and trade elasticities are considered to be fixed values in the reduced-form metamodel:

$$\begin{aligned}\hat{z} =& (\alpha_0 + \sum_j \beta_j \bar{\theta}_j + \sum_j \sum_{j'=j+1} \gamma_{jj'} \bar{\theta}_j \bar{\theta}_{j'} + \sum_j \sum_j \gamma_{jj} \bar{\theta}_j^2) \\ &+ \sum_i (\alpha_i + \sum_j \beta_{ij} \bar{\theta}_j) tp_i \\ &+ \sum_i \sum_{i'=i+1} \alpha_{ii'} tp_i tp_{i'} + \sum_i \alpha_{ii} tp_i^2,\end{aligned} \tag{11}$$

where the general settings remain the same with the only exception that in the reduced-form meta-models θ's are fixed values and thus they are grouped into other coefficients. In this case, they are grouped into both the constant and the coefficients of the main technical progress shock effects.

$$\begin{aligned}y = \alpha_0 &+ \sum_i \alpha_i tp_i + \sum_j \beta_j \theta_j \\ &+ \sum_i \sum_{i'=i+1} \alpha_{ii'} tp_i tp_{i'} + \sum_i \sum_j \beta_{ij} tp_i \theta_j + \sum_j \sum_{j'=j+1} \gamma_{jj'} \theta_j \theta_{j'} \\ &+ \sum_i \alpha_{ii} tp_i^2 + \sum_j \gamma_{jj} \theta_j^2 + \epsilon\end{aligned} \tag{12}$$

$$
\begin{aligned}
\hat{y} =&(\alpha_0 + \sum_j \beta_j \bar{\theta}_j + \sum_j \sum_{j'=j+1} \gamma_{jj'} \bar{\theta}_j \bar{\theta}_{j'} + \sum_j \sum_j \gamma_{jj} \bar{\theta}_j^2) \\
&+ \sum_i (\alpha_i + \sum_j \beta_{ij} \bar{\theta}_j) tp_i \\
&+ \sum_i \sum_{i'=i+1} \alpha_{ii'} tp_i tp_{i'} + \sum_i \alpha_{ii} tp_i^2,
\end{aligned} \tag{13}
$$

We want to point out that the reduced-form meta-models are constructed not estimated but still we can validate them by using the reduced-form meta-models to predict the outputs based on all the observations and determine their performance via the two aforementioned criteria *rmse* and *mare*. In addition, we are interested in the marginal effects of technical progress shocks on the outputs which can be calculated from the following equation:

$$
\begin{aligned}
\frac{\partial \hat{z}}{\partial tp_i} =&\alpha_i + \sum_j \beta_{ij} \bar{\theta}_j \\
&+ \sum_{i' \neq i} \alpha_{ii'} tp_{i'} + 2 * \alpha_{ii} tp_i
\end{aligned} \tag{14}
$$

It is noticeable that the term $\sum_j \beta_{ij} \bar{\theta}_j$ represents the impact from the elasticity parameters which can also be understood as the impact from the CGE models. Therefore, we would like to find out whether there are significant variations within it and compare the magnitude of it with the magnitudes of other terms which affect the marginal effects in order to test to what extent the elasticity parameters affect the marginal effects.

4 Results

4.1 Fitting and Validation

4.1.1 General Meta-model

We apply the method to three outputs $Z1$ (welfare of small-scale farmers), $Z2$ (poverty reduction) and $Z5$ (urban consumer welfare). The sample of simulation inputs remains unchanged for the three cases.

Firstly, let's have a look at the fitting performance. The R_{adj}^2 are 0.998, 0.9998 and 0.998 respectively, which means that the fitting works quite well for the three outputs.

Secondly, let's have a look at the validation performance which is summarized in Table 2.

Although there is generally not a lower threshold for *rmse* but in practice we can compare the statistic with the original range of the outputs. For example, if the output

	In-Sample rmse	In-Sample mare	Out-of-Sample rmse	Out-of-Sample mare
Z1	0.821	0.0035	0.808	0.0034
Z2	0.008	0.0004	0.012	0.0014
Z5	2.792	0.0027	2.750	0.0051

Table 2: Validation Performance

varies from 0 to 1000 and the *rmse* is 0.7 we may conclude that we have a good prediction but if the range of the output is from 0 to 1 and the *rmse* is still 0.7, we may conclude that the prediction works not so well. In our data, the ranges of $Z1$, $Z2$ and $Z5$ are [1061.125,1282.656], [88.500,91.584] and [4460.579,5373.743] respectively. We can easily calculate the ratio of the *rmse* and the averaged values for the three outputs which are listed in Table 3. The fact that these ratios are so small enables us to conclude that our meta-model has done an excellent job in prediction. While for *mare*, a recommended lower threshold is 0.1 and we can easily notice that the values for both cases are much lower than this threshold, which gives us another message that our meta-model has behaved well in modeling the relationshipbetween the inputs and outputs. Besides, the fact that the in-sample *rmse* and the out-of-sample *rmse* are similar also represents that the model is well-constructed.

	In-Sample Ratio	Out-of-Sample Ratio
Z1	0.0007	0.0007
Z2	0.00008	0.00013
Z5	0.0006	0.0006

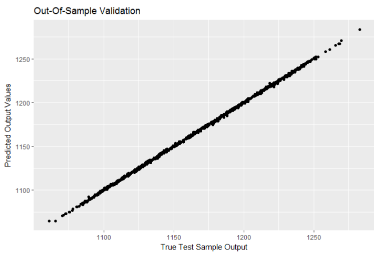
Table 3: Ratio of RMSE and Averaged Outputs

Besides, we can have a look at Figure 4 which shows the predicted outputs versus simulation outputs for both the training sample and the test sample. The fact that there are not clear deviations of the predicted values from the true values (almost a perfect fit) leads us again to the conclusion that the meta-model works well in modeling the behavior of the simulation model (the Senegal-CGE model) and can be used as a surrogate of the simulation model in the following analysis. In other words, if we intend to achieve other research purposes with the help of the simulation model, we can now use the meta-model to replace the simulation model. This will not only make the analysis faster and easier but will also make some previously-not-feasible approach feasible, such as the Bayesian Model Selection approach.

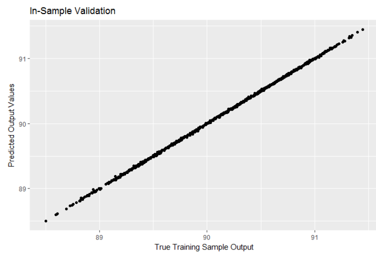
4.1.2 Reduced-form Meta-model

Since the general meta-model has been validated, we can furthermore validate all the reduced-form metamodels by using each reduced-form meta-model to predict the outputs

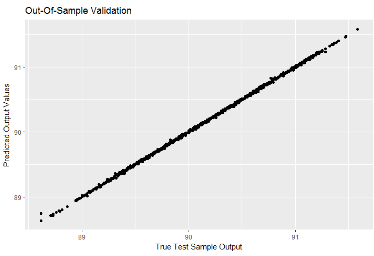
(a) In-Sample(Z1)



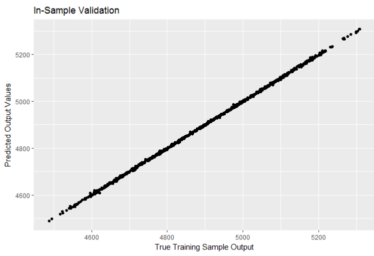
(b) Out-of-Sample(Z1)



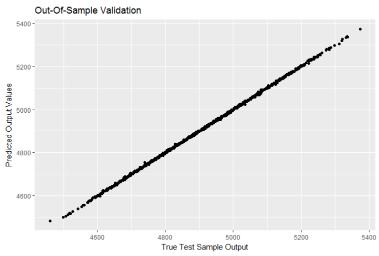
(c) In-Sample(Z2)



(d) Out-of-Sample(Z2)



(e) In-Sample(Z5)



(f) Out-of-Sample(Z5)

Figure 4: Prediction Performance

based on each set of technical progress shocks. Moreover, we can compare the *rmse* and *mare* of the predictions from all reduced-form meta-models in order to quantify the performance. We want to point out that in the generation of reduced-form meta-models we don't distinguish between training set and test set.

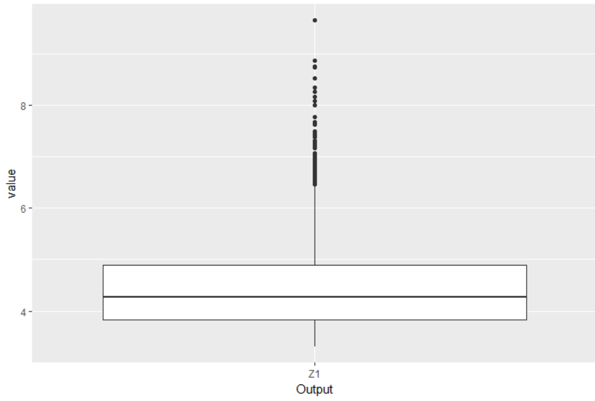


Figure 5: RMSE of Reduced-form Meta-models for Z1

With respect to $Z1$, the averaged *rmse* and *mare* are 4.477 and 0.014. (Table 4) The ratio of the averaged *rmse* and the correspongding averaged outputs is 0.0038. In comparison with the validation performance of the general meta-model, we can readily see that the reduced-form meta-models have much larger *rmse* and *mare* than that of the general meta-model meaning that the general meta-model has better prediction capability. However, we could have foreseen this result because we have constructed the reduced-form meta-models on the basis of the general meta-model such that they have less explanatory variables which decrease their capacity of prediction. Nevertheless, we can still come to the conclusion that the reduce-form meta-models are accepted based on the validation results. Figure 5 displays the distribution of *rmse* which proves that the prediction performance of reduced-form meta-models are in general still quite good.

	mean rmse	mean mare
Z1	4.477	0.014
Z2	0.052	0.002
Z5	14.611	0.012

Table 4: Reduced-form Meta-model Performance

(a) RMSE for Z2

(b) RMSE for Z2

Figure 6: RMSE of Reduced-form Meta-models for Z2 and Z5

With respect to $Z2$ and $Z5$, we arrive at similar conclusions which can also be shown by Table 4 and Figure 6 that the reduced-form meta-models work well in predicting the outputs meaning that we can use the reduced-form meta-models to perform the analysis in the next step.

4.2 Testing the Impact of Elasticity Parameters

In practice, we find out that $\alpha_i + \sum_j \beta_{ij}\bar{\theta}_j$ is much larger than $\sum_{i' \neq i} \alpha_{ii'} tp_{i'} + 2 * \alpha_{ii} tp_i$, moreover, the coefficients $\alpha_{ii'}$ and α_{ii} are statistically insignificant, therefore, to simplify the calculation, we measure the marginal effects of technical progress shocks only by $\alpha_i + \sum_j \beta_{ij}\bar{\theta}_j$. As α_i remains fixed with repect to all the reduced-form meta-models, the variability of the marginal effect comes only from the variability of $\sum_j \beta_{ij}\bar{\theta}_j$, which is also interpreted as the impact from each unique CGE models. Figure 7,9 and 11 show the distribution of the sector-specific marginal effects of technical progress shocks on the outputs $Z1$, $Z2$ and $Z5$ respectively and figure 8, 10 and 12 are the corresponding histogram of the sector-specific marginal effects. We use two measures to determine if there is variability in the sector-specific marginal effects, namely, the interquartile range and standard deviation. By definition, the interquartile range of a box-whisker plot includes the middle 50% of the data and the larger the interquartile range, the more variable the data set is. From figure 7,9 and 11 we can come to the conclusion that for all the three outputs, the interquartile ranges of marginal effects of every sector are extremely narrow meaning that the sector-specific marginal effects are quite stable in this case, however, we can see that for different sectors, the marginal effects differ obviously, which means that sectors have quite distinct impacts on the outputs. Moreover, from figure 8, 10 and 12 we can detect that the marginal effects of each sector for all the three outputs is distributed normally or approximately normally, and thus the 95% of the distribution is within two standard deviations from the mean. In our case, the standard deviation of the marginal effects of each sector for all the three outputs are relatively low, mostly varies between 0.5 and 2, therefore, we can again conclude that the marginal effects are not variable. The message sent by these two statistics is that there is no signifcant variability in the marginal effects which is equivalent to the message that in this application there is no significant impact on the outputs from the various combinations of elasticity parameters of the CGE models.

The elasticity parameters are very crutial in CGE modeling (Lofgren et al. (2002) and Fan (2008)). Some of them need to be assumed while some of them need complex econometric models for estimation, thus in practice, researchers either extract them from literature based on one reason or the other or invest much in estimation. Afterward, for the sake of comparison, they usually perform the sensitivity analysis by giving the parameters high and low values so as to compare the results and draw conclusions. In our demonstration, we introduce the meta-modeling technique which is an intermediate step towards the application of Bayesian model selection method that could solve the CGE modeling problem from a totally new perspective.

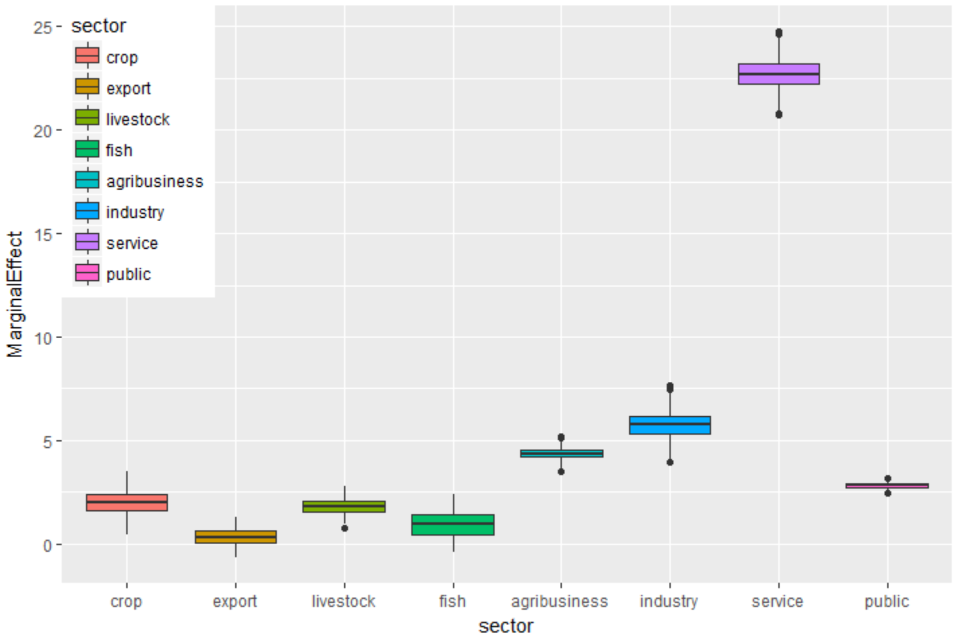
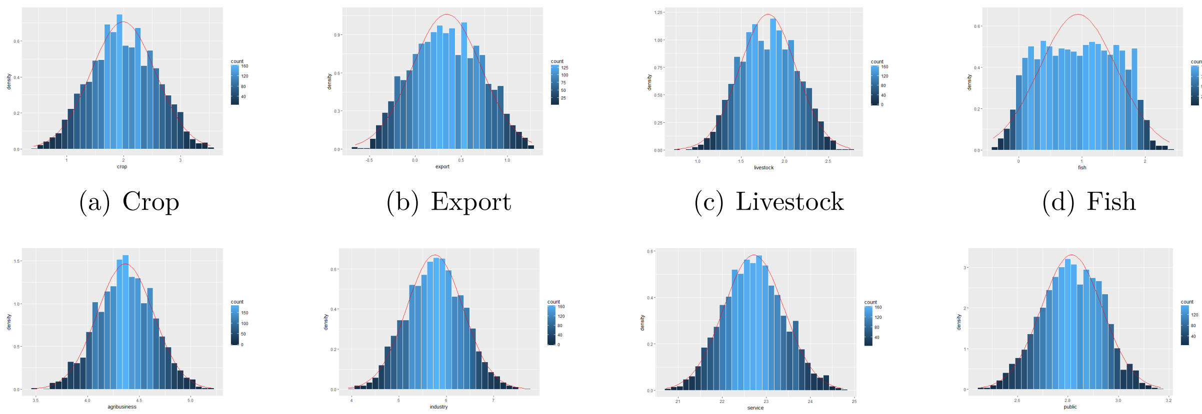


Figure 7: Marginal Effects of Technical Progress Shocks on Output Z1



(a) Crop (b) Export (c) Livestock (d) Fish

(e) Agribusiness (f) Industry (g) Service (h) Public

Figure 8: Histogram of Marginal Effects of Technical Progress Shocks on Output Z1

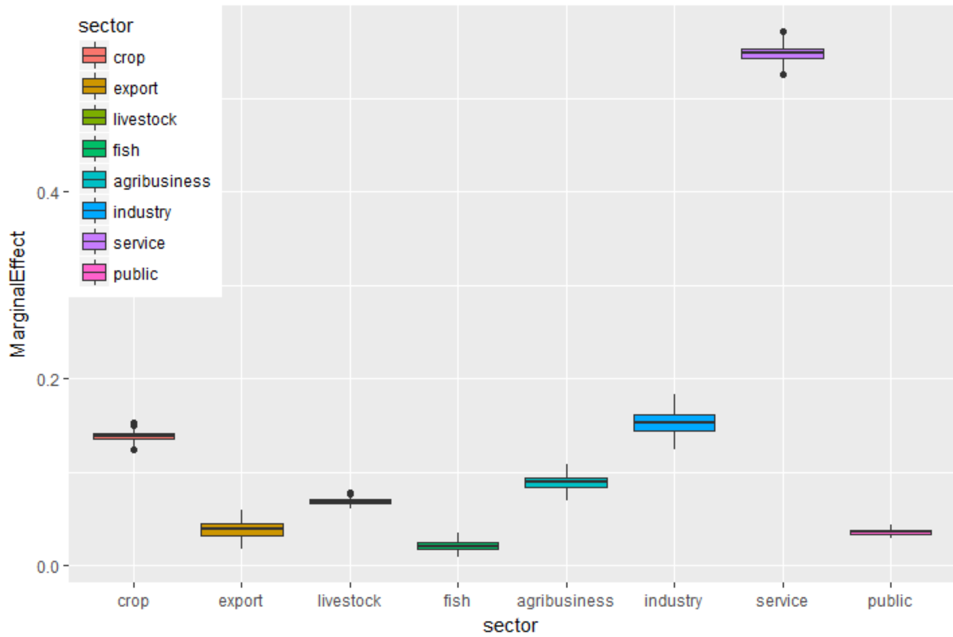
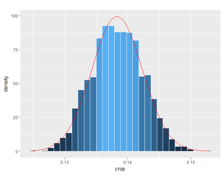
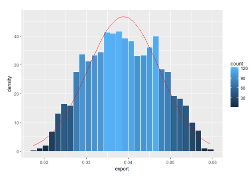
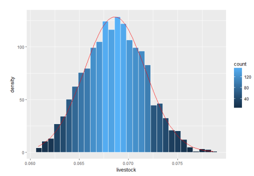
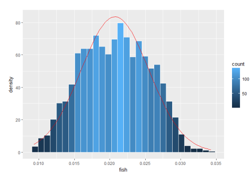


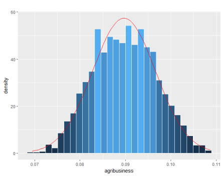
Figure 9: Marginal Effects of Technical Progress Shocks on Output Z2

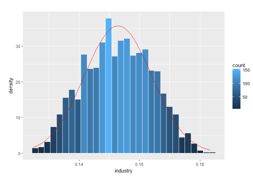

(a) Crop

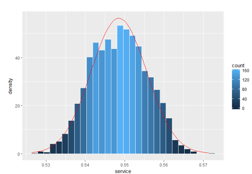

(b) Export


(c) Livestock


(d) Fish


(e) Agribusiness


(f) Industry


(g) Service

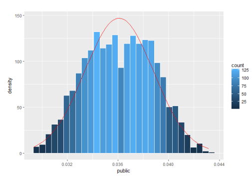

(h) Public

Figure 10: Histogram of Marginal Effects of Technical Progress Shocks on Output Z2

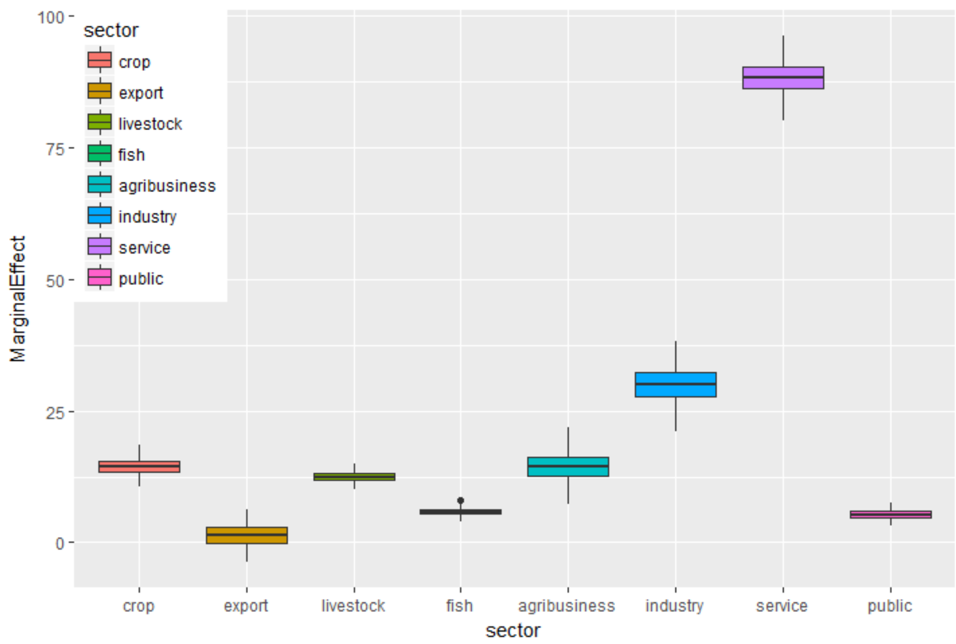
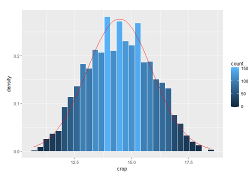
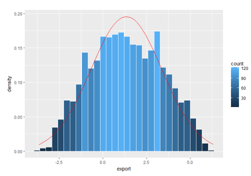


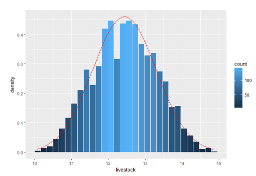
Figure 11: Marginal Effects of Technical Progress Shocks on Output Z5



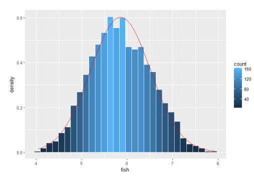
(a) Crop



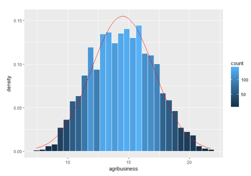
(b) Export



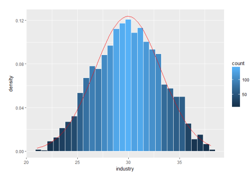
(c) Livestock



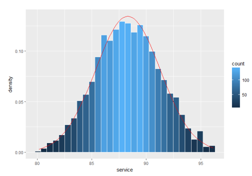
(d) Fish



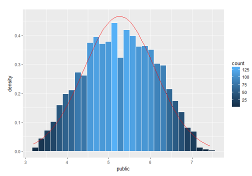
(e) Agribusiness



(f) Industry



(g) Service



(h) Public

Figure 12: Histogram of Marginal Effects of Technical Progress Shocks on Output Z5

5 Conclusion

A highly discussed and criticized aspect of CGE modeling is the elasticity parameters, researchers usually extract them from literature because of either the complex and tedious estimation process or the impossibility of estimation and use them directly in the simulations. Moreover, due to the complexity of CGE models, it is difficult to combine them with other approaches such as Bayseian Model Selection or incorporate them into other models such as the decision-making model. Therefore, we apply the meta-modeling methodology to generate valid surrogates of the CGE models because we are looking forward to deal with the CGE modeling issue from another angle. In this paper, we have shown how to apply the meta-modeling technique and proved that the generated meta-models are indeed valid surrogate models of the CGE models, which makes it sensible to use them instead of the CGE models in other analyses. Although we have not detected the impact on the outputs from the various specific elasticity parameter setting of the CGE models, the application of meta-modeling technique still opens the door to many possibilities. In addition, we will include closure rules into the meta-modeling technique in the future.

References

Arndt, Channing, Sherman Robinson, and Finn Tarp, "Parameter estimation for a computable general equilibrium model: a maximum entropy approach," *Economic Modelling*, 2002, *19* (3), 375–398.

Barton, Russell R, "Tutorial: simulation metamodeling," in "Proceedings of the 2015 Winter Simulation Conference" IEEE Press 2015, pp. 1765–1779.

Blanning, Robert W, "The sources and uses of sensitivity information," *Interfaces*, 1974, *4* (4), 32–38.

_ , "The construction and implementation of metamodels," *simulation*, 1975, *24* (6), 177–184.

Bourguignon, François, *The impact of economic policies on poverty and income distribution: evaluation techniques and tools*, World Bank Publications, 2003.

Box, G. E. P. and K. B. Wilson, "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1951, *13* (1), 1–45.

Chen, Victoria CP, Kwok-Leung Tsui, Russell R Barton, and Martin Meckesheimer, "A review on design, modeling and applications of computer experiments," *IIE transactions*, 2006, *38* (4), 273–291.

Cressie, Noel and Ngai H Chan, "Spatial modeling of regional variables," *Journal of the American Statistical Association*, 1989, *84* (406), 393–401.

Eriksson, L, E Johansson, N Kettaneh-Wold, C Wikström, and S Wold, "Design of experiments," *Principles and Applications, Learn ways AB, Stockholm*, 2000.

Fan, Shenggen, *Public expenditures, growth, and poverty: Lessons from developing countries*, Vol. 51, Intl Food Policy Res Inst, 2008.

Giunta, Anthony A, Steven F Wojtkiewicz, Michael S Eldred et al., "Overview of modern design of experiments methods for computational simulations," in "Proceedings of the 41st AIAA aerospace sciences meeting and exhibit, AIAA-2003-0649" 2003.

Hazledine, Tim et al., "A critique of computable general equilibrium models for trade policy analysis," Technical Report 1992.

Kleijnen, Jack PC, "A comment on Blanning's Metamodel for sensitivity analysis: the regression metamodel in simulation," *Interfaces*, 1975, *5* (3), 21–23.

_ , *Design and analysis of simulation experiments*, Vol. 20, Springer, 2008.

_ **and Charles R Standridge**, "Experimental design and regression analysis in simulation: An FMS case study," *European Journal of Operational Research*, 1988, *33* (3), 257–261.

_ **and Robert G Sargent**, "A methodology for fitting and validating metamodels in simulation," *European Journal of Operational Research*, 2000, *120* (1), 14–29.

_ **, Susan M Sanchez, Thomas W Lucas, and Thomas M Cioppa**, "State-of-the-art review: a users guide to the brave new world of designing simulation experiments," *INFORMS Journal on Computing*, 2005, *17* (3), 263–289.

Lofgren, Hans and Sherman Robinson, "Public spending, growth, and poverty alleviation in Sub-Saharan Africa: A dynamic general equilibrium analysis," *Public expenditures, growth, and poverty: lessons from developing countries*, 2008.

_ **, Rebecca Lee Harris, and Sherman Robinson**, *A standard computable general equilibrium (CGE) model in GAMS*, Vol. 5, Intl Food Policy Res Inst, 2002.

McKay, M. D., R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 1979, *21* (2), 239–245.

Noordegraaf, Antonie Vonk, Mirjam Nielen, and Jack PC Kleijnen, "Sensitivity analysis by experimental design and metamodelling: Case study on simulation in national animal disease control," *European Journal of Operational Research*, 2003, *146* (3), 433–443.

Simpson, Timothy W, Dennis KJ Lin, and Wei Chen, "Sampling strategies for computer experiments: design and analysis," *International Journal of Reliability and Applications*, 2001, *2* (3), 209–240.

Srivastava, A, Kurt Hacker, Kemper Lewis, and TW Simpson, "A method for using legacy data for metamodel-based design of large-scale systems," *Structural and Multidisciplinary Optimization*, 2004, *28* (2-3), 146–155.

Stocki, Rafal, "A method to improve design reliability using optimal Latin hypercube sampling," *Computer Assisted Mechanics and Engineering Sciences*, 2005, *12* (4), 393.

Wang, Chunhua, "A Dynamic Stochastic Frontier Production Model with Time-varying Efficiency: Comment," *Applied Economics Letters*, 2007, *14* (6), 415–417.